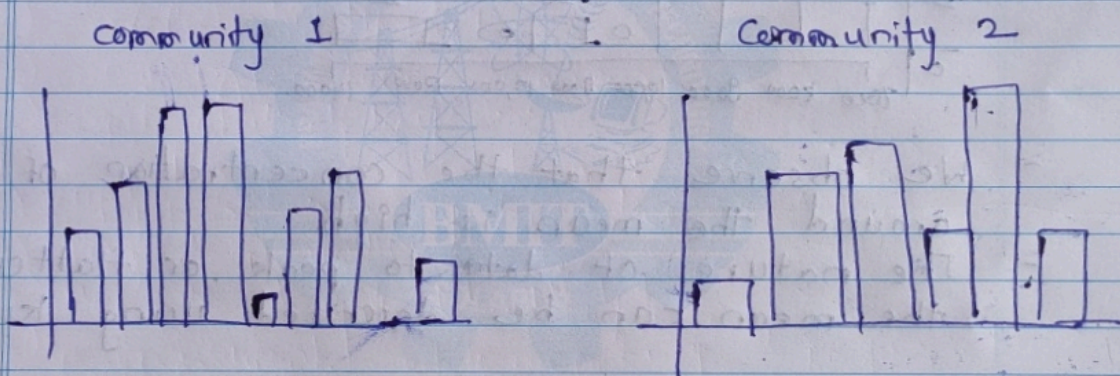


* Measure of shape *

- To understand the distribution of data statistically we use the measures of the shape such as skewness and kurtosis to describe the data set better.

Histogram : To represent the distribution of numeric data graphically we use a histogram.

(i) skewness



consider, for given data.

Mean of community 1 = 8683.4
Median " " 1 = 8357

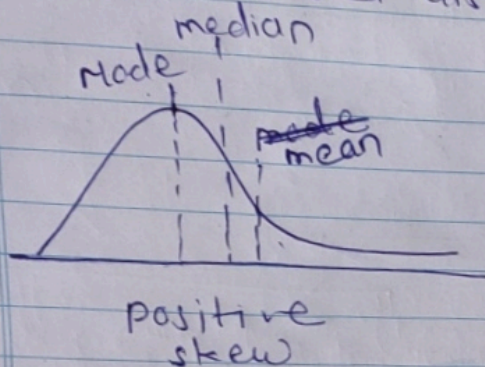
Mean of community 2 = 8225.6
Median " " 2 = 8213

- If there is an unusually large value in the dataset, the mean becomes greater than the median, this indicates right skewed. On other hand, if there are ~~unusually~~ unusually small values, the mean becomes less than the median, this indicates left skewed data.

Right, Left & Normal Distribution

B.M.I.T.

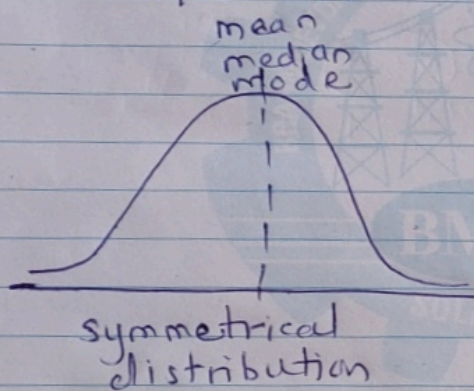
① Right skewed distribution



$$\boxed{\text{mode} < \text{median} < \text{mean}}$$

- e.g.:
- ① Wealth distribution
 - ② length of comments in video

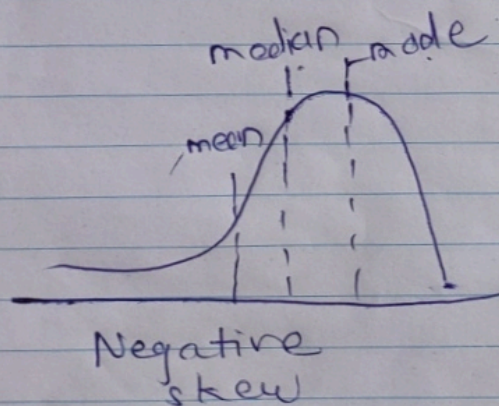
② Normal / Gaussian Distribution



$$\boxed{\text{mean} = \text{median} = \text{mode}}$$

- e.g.:
- ① Age
 - ② Weight
 - ③ Height

③ Left skewed distribution



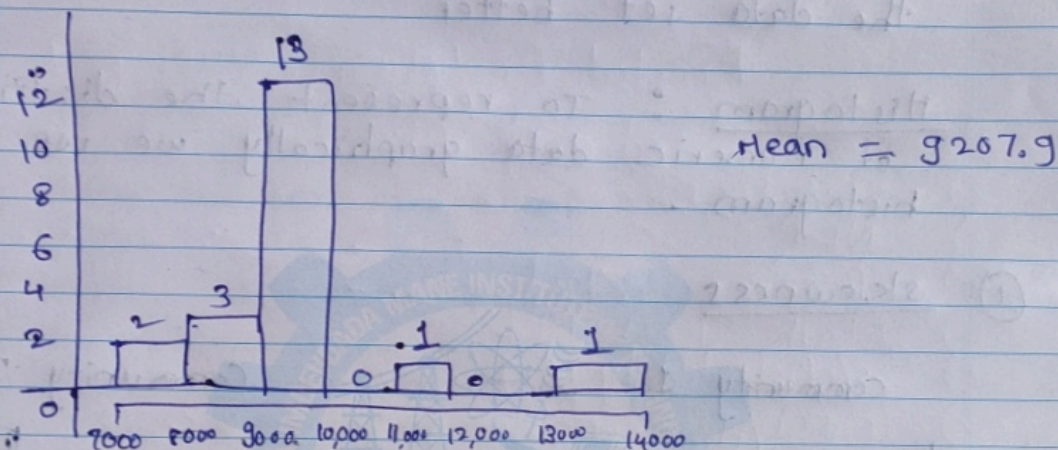
$$\boxed{\text{mean} > \text{median} > \text{mode}}$$

- e.g.:
- ① Life span of human being

$$\text{skewness} = 3 * (\text{Mean} - \text{Median}) / \text{std deviation}$$

② kurtosis :-

E.g



- We observe that the concentration of data around the mean is high.
- The nature of data to peak or flatten about the mean can be described using kurtosis.
- kurtosis helps us identify whether the data spread around the mean is high or not.
- A higher value of kurtosis indicates that data is largely centred around the mean, while a lower of kurtosis indicates that data is not centred around mean.

$$\text{kurtosis Formula} = \frac{\sum_{i=1}^N \left(\frac{x_i - \bar{x}}{s} \right)^4}{N}$$

- The kurtosis for dataset above is 7.844213, which

- indicates a high peak.
- A value of kurtosis above 3 indicates a peak, while that below 3 indicates flatness

Ques: A distribution having high positive kurtosis value will have - - - - - when ordered

- ☒ i) Less number of values at the tail end of data
- ☐ ii) More no. of values at the tail end of data
- ☒ iii) More no. of values at the center of the data
- ☐ iv) Less no. of values